

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

POD Quarterly: The Journal of the Professional
and Organizational Development Network in
Higher Education

Professional and Organizational Development
Network in Higher Education

Winter 1980

The Verbal Structure of Teacher Questions: Its Impact on Class Discussion

John D. W. Andrews

Follow this and additional works at: <https://digitalcommons.unl.edu/podqtrly>



Part of the [Higher Education Administration Commons](#)

Andrews, John D. W., "The Verbal Structure of Teacher Questions: Its Impact on Class Discussion" (1980).
POD Quarterly: The Journal of the Professional and Organizational Development Network in Higher Education. 32.

<https://digitalcommons.unl.edu/podqtrly/32>

This Article is brought to you for free and open access by the Professional and Organizational Development Network in Higher Education at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in POD Quarterly: The Journal of the Professional and Organizational Development Network in Higher Education by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

The Verbal Structure of Teacher Questions: Its Impact on Class Discussion

JOHN D. W. ANDREWS

Improving discussion participation is one of the most widespread concerns of college instructors. Like most consultants, I often make suggestions based on my own experience and on hunches formed from observing a variety of teachers at work. Yet I have felt the need to test these hunches, and to generate solutions to the problem that are based on objective, systematic evidence. To take a step in this direction has been the purpose of the research project presented here.

This study deals with the contribution made by the *verbal structure* of questions to the form and quality of discussions.¹ This is a fruitful place to start the analysis of discussions, since it is a factor which can be easily abstracted and studied apart from the interwoven context of classroom process. In addition to its "studyability," this sphere should produce useful information since the phrasing of questions is largely under the control of the instructor, and thus provides leverage for the person who wants to improve his or her teaching.² This is so in two senses: first, it is the instructor, rather than students, who poses most questions; and second, questioning style—unlike many expressive, personality-linked characteristics—can be altered by a deliberate effort.

¹ An aspect of this project formed the basis for a presentation at the 1979 POD Annual Conference. It dealt with one particular question form, the "Focal Question," which will be discussed later in this paper.

² My appreciation to Sondra Napell, a higher education consultant, who first directed my attention to question form. She uses this approach in helping instructors to analyze written transcripts of their own classroom questions.

POD Quarterly, Vol. 2, Nos. 3 & 4 (Fall/Winter 1980)

It is useful at this point to reformulate our main objective as a pair of hypotheses or questions: *Does* the verbal form given to instructor-initiated questions influence student response, independent of other variables and the context of the class as a whole? And if it does, *what* question forms are the most fruitful? To assemble evidence on these points, two things are required: first, a clear, measurable definition of "good discussions"; and second, a way of classifying questions in order to distinguish what kinds have what effects. The procedure I used was to first assess the quality of discussions, recorded on videotape, from a number of classes in the humanities disciplines; and then to examine the questions which preceded the "good" and "poor" discussions in search of patterns and common themes.

But how can one measure discussion quality? The approach I took was to concentrate on the nature of interaction among instructor and students, an emphasis which corresponds well with what many instructors tell me they want in their discussions. A good discussion is one in which each point raised elicits a variety of student responses. In such discussions, a large portion of the class is active, rather than only a minority. And good discussions have momentum—students will continue interacting for some time without the need of further questions or prompting. Finally, in good discussions students engage each other in conversation; they have abandoned the pattern in which each student contribution must be followed by a comment of some kind from the instructor. These various characteristics, all reflecting the extent of student response, can be encompassed by the inelegant but serviceable term, "mileage."³

THEORETICAL CONSIDERATIONS

In searching for the ingredients of good discussions, I was guided by several hunches and hypotheses. The distinction drawn by Guilford (1958) between convergent and divergent thinking has seemed to me a useful guideline for forming discussion questions. Divergent-

³ Of course, the *caveat* should be added here that because of the many differences among instructors' goals, values, and personal styles, this is not everyone's idea of a good discussion. Some may not feel the interactive pattern described here is a fruitful one. And this method says nothing about the content of questions or the quality of thought expressed by the participating students. The purpose of this research is to draw, as clearly as possible, a set of means-end relationships which can be used by instructors to reach certain goals when these are considered desirable.

thinking questions, for which there can be a number of "correct" or discussible answers, are more likely to generate continuing discussion. With convergent questions, by contrast, participation may be inhibited because of fears about being wrong, and the dialogue will probably grind to a halt when "the" answer sought by the instructor is produced. For example, a question like:

How else would you end this story, if you couldn't have him reach some sort of point of divine revelation . . . all we're allowed to do is rewrite the ending. Tolstoy won't let us rewrite the rest.

seems more likely to generate extended discussion than:

What's the main message here? What's Hemingway trying to tell us in this story?

Higher level questions,⁴ characterized by the intellectual operations of analysis, synthesis, or evaluation, seem more likely to produce extended discussion as students become engaged with the complex issues involved. By contrast, lower level questions, drawing on memorization, comprehension, or application, can generally be dealt with more briefly. For example:

What about the function of Laertes . . . what is his role . . . for example, could Laertes be left out of the work and would it still be *Hamlet*?

will probably be explored at some length, while this question:

Okay, men move out to work . . . where did they used to work?

requires only a short answer.⁵

Another lead emerged from the frustrations of trying to decipher the impact of certain questions, in both consulting and research contexts. When one listens closely to classroom questions, or even better sees them transcribed (and thus compares them with written expression), it is immediately, staggeringly obvious that many of them are very poorly constructed and phrased. Often the instructor

⁴ This conceptualization is drawn from the taxonomy developed by Bloom *et al* (1956).

⁵ The two distinctions drawn so far (convergent versus divergent and higher versus lower level) may seem to overlap considerably. Yet I believe that they are conceptually distinct; the comprehensive typology to be presented later in this article will provide examples of all four combinations of the two dimensions.

uses unclear language, tries a couple of versions of the question in succession, interrupts himself or herself to toss in a verbal footnote or background information, oscillates between memory-oriented and analytical questions, or even asks several unrelated questions in the same breath. For example, consider:

Does that seem to be what Tolstoy is condemning him for mainly? . . . What seems to be the main problem, according to Tolstoy? At the end of the story, we have a religious solution; a couple of you earlier said you didn't think that quite fit with the rest of the story . . . do you still feel that way after discussing it? What's the problem there?

as opposed to:

And did the reform work? How was corruption under the corregidor?

These questions gave me fits when I tried to set up clear and consistent categories; I accepted this as the lot of a content-analysis researcher, but it also struck me that my reaction would probably be duplicated in the minds of students as they tried to grasp what the instructor wanted. This line of thinking produced a third hunch: that questions which have multiple, inconsistently structured subparts will yield less discussion mileage than clear, consistent questions because student confusion will be an obstacle to participation.

A fourth issue arose from consultation work with instructors who, in their eagerness to create open discussions, seem to have overshot and produced vagueness and amorphousness instead; the result is often an "escape from freedom" reaction in students, whose confusion and anxiety lead them to retreat from participation. For example, a question such as:

How did you like the play?

while apparently offering many openings for students involvement, can be less productive than a more structured question like:

What are some similarities or differences you note in comparing Hamlet and Laertes?

A final hunch concerned "wait-time"—the post-question pause in which the instructor anxiously anticipates a student response of some sort. As has been pointed out by Napell (1976), not allowing

enough wait-time is a common teaching mistake. This usually happens because the pause is a fertile breeding ground for the instructor's fantasies about the shortcomings of the question, his or her teaching weaknesses, and so on. As one professor put it to me during a consultation session, "Silence is failure!" Unfortunately, leaping in too quickly to amend the question, or ask a "better" one, usually spoils the original question and muddies the water by giving the students too much to think about at once.

I often suggest that instructors learn to wait longer after questions, and support this point by enumerating the mental operations which a student must go through in order to prepare a response. It is necessary to grasp the question; decide what expectations the instructor holds; search one's memory storage for relevant ideas or information; select a response; and overcome any inhibitions about speaking in a group or making a mistake. I also point out that the more complex or difficult the question, the more extensive the student's invisible mental activity must be. Thus, if one expects students to deal with challenging questions, one should be prepared to wait long enough for them to work out useful ideas in response. It is only the trivia which can be reacted to immediately.

FORMAL HYPOTHESES

The foregoing considerations led to a set of research hypotheses focused in four main areas: (I) relationships among various indices of discussion quality built on the interactional characteristics described earlier; (II) the extent of discussion "mileage" associated with the various question characteristics just discussed; (III) the relationship of question complexity to wait-time; and (IV) a holistic taxonomy of question "species."

Hypothesis I:

It will be possible to develop an internally consistent operational definition of discussion quality, based on five indices drawn from the general definition stated earlier. The five indices are:

- a) Number of Student Statements (NSS): The number of individual student contributions, without regard for the number of speakers, which follows a given question.
- b) Number of Students (NS): The number of individuals who are active in the conversation following a given question.
- c) Number of Student-Follows-Student Sequences (NS-S): The num-

ber of instances, following a given question, in which one student verbalization is followed immediately by another, without an intervening verbalization by the instructor.

d) Student Talk Time (STT): The duration, in number of seconds, of all student talk following a given question.

e) Percent Student Talk (Versus Teacher Talk) (%S): The number of separate student verbalizations divided by the total number of verbalizations (by instructor and students), which follows a given question.

It is predicted that these five indices will be positively intercorrelated, thus permitting the formulation of a single unified index of discussion participation.

Hypothesis II:

That certain question characteristics will be associated with higher or lower levels of discussion participation.

a) That divergent questions will yield greater participation than convergent questions.

b) That questions calling for higher levels of thinking will produce more participation than lower-level questions.

c) That clear, consistent questions which raise a single point and call for one type of thinking from students will lead to greater amounts of discussion than questions which have inconsistent subparts and thus set contradictory expectations.

d) That questions which provide an orientation or focus for student responses will be more effective in generating discussion than questions that are lacking in guidelines or direction.

Hypothesis III:

That questions requiring considerable interpretation and/or complex thought will be followed by longer "wait-time" pauses than simpler, more obvious questions.

a) That questions involving higher levels of thinking will be followed by longer pauses than those involving lower levels of thinking.

b) That vague, unstructured questions will be followed by longer pauses than those which provide structure and orientation.

c) That questions which have inconsistent subparts will be followed by longer pauses than clear, straightforward questions.

Hypotheses IIIa-c duplicate the variables thought likely to affect discussion mileage (as presented in Hypothesis II), with one excep-

tion: the convergent versus divergent dimension. The omission is deliberate, because the basis of prediction concerning this variable is not clear. One might expect divergent questions to lead to longer wait-times because there are more possible answers to consider. On the other hand, convergent questions (at an equivalent level of complexity) require more careful evaluation of the *correctness* of the response; and there may also be more inhibiting anxiety about being wrong. These reciprocally operating factors seem most likely to cancel each other out.

Hypothesis IV:

This is more a heuristic than a hypothesis in the strict sense. It arises from the impression that beyond the single variables enumerated above, questions can be classified by a more encompassing, holistic taxonomy. That is, certain recurrent themes may yield distinct question "species," through combining single characteristics into more complex patterns. I pursued Hypothesis IV, not by a direct quantitative test, but by an impressionistic scanning of the data followed by the gradual crystallization of impressions; these were then formalized, cross-validated, and checked against discussion quality criteria.

PROCEDURES

The research was based on seven 45-minute videotapes of T.A.-led discussion sections in the humanities disciplines (Literature, History and Humanities) at the University of California, San Diego. Most of the students in these classes were freshmen, and the T.A.s had varying degrees of experience in teaching (ranging from 2 to 6 academic quarters).

On each tape, two successive 10-minute segments were marked off and designated as Samples 1 and 2. Sample 1 was divided into subsamples 1a and 1b, and Sample 1a was used to conduct exploratory, inductive analysis. Sample 1b served to further refine and test these procedures, and Sample 2 served as a final cross-validation of the coding system. In the data tables below, Samples 1 and 2 combined; any discrepancies between the two samples are noted in the text. All categorizations, other than purely mechanical counting (which involved primarily the discussion quality and wait-time measurements), were written up in a formal coding manual which was in turn applied by a second rater. After preliminary discussions

(using Sample 1a), the manual was refined and adequate levels of inter-rater agreement were achieved. To ensure that preconceptions about results would not influence the classification procedure, the coding of questions in Sample 2 was done by one rater, who did not listen to the student-response portions of the tapes. All counting of "mileage" scores was done by the second rater, without knowledge of the questions.

The definition of a teacher question, applied to 14 questions, yielded 100% agreement. Regarding the question classification categories,⁶ the overall level of agreement for Sample 1 was 80% (based on 146 separate coding decisions), and for Sample 2 it was 85% (based on 90 coding decisions). This level was generally constant across categories. In the last reliability check (Sample 2) only one of ten categories was below 80% agreement (at 60%), and many were at the 100% agreement level. At this point, remaining differences were resolved by discussion in order to arrive at final classifications for all questions.

FINDINGS

The data concerning Hypothesis I not only test a specific prediction, but are basic to subsequent analyses since the index of discussion quality is the criterion for testing Hypotheses II and III. Results for Sample 1 are shown in Table 1, in the form of Pearson correlations among the five variables described previously. There is one

TABLE 1
INTERCORRELATIONS AMONG FIVE INDICES OF DISCUSSION QUALITY

	NSS	NS	STT	NS-S	%S
NSS	—	+.89	+.80	+.93	+.06
NS	+.89	—	+.68	+.77	-.01
STT	+.80	+.68	—	+.74	+.05
NS-S	+.93	+.77	+.74	—	+.18
%S	+.06	-.01	+.05	+.18	—

NOTE: N = 68 throughout (Sample 1 only). Correlation coefficients of $\pm .68$ and larger are statistically significant at $P < .001$. Those of $\pm .18$ or less are not statistically significant at $P < .05$.

⁶ The coding system was constructed so that each coding decision was a dichotomous, "either-or" or "present-absent" decision. Each rater made a decision about each question in a given sample; agreement is defined as the number of identical decisions divided by the number of questions in the sample. (More detailed technical information, including a copy of the coding manual, can be obtained by writing the author.)

broad pattern which encompasses four indices, and a single unrelated variable. The methods of measuring student activity in a direct, linear way all show strong covariation, while %S, which is a *ratio* of student talk to teacher talk, is orthogonal to the others. This pattern can be pinpointed even more clearly by computing the same correlation matrix separately for each instructor. For the four highly correlated variables, there is little variation from instructor to instructor, while the correlations between %S and the other four vary widely—from $+.63$ to $-.83$.

These findings suggest several reasons for using the four intercorrelated measures as our index of discussion mileage. They are methodologically similar, all being based on student activity (rather than a student-teacher ratio); they hang together statistically; and they represent differing but related facets of the definition offered earlier for a "good discussion." Finally, these four measures show considerable internal consistency, whereas the %S score does not.

Because the intercorrelations among the four clustered measures are so high in Sample 1, a single one of them is used as a representative index in Sample 2 and subsequent data analysis. The NSS score seems the best choice for several reasons: it has the highest set of intercorrelations with the other three measures; it is conceptually central in that it deals with the overall scope of student response; and it shows a wide range of variation. The use of a single quality measure greatly simplifies analytical procedures and eliminates problems of combining measures appropriately.

Table 2 presents data used to test Hypotheses IIa-c (Hypothesis II d will be considered at a later point). Clear-cut confirmations appear regarding the first two of these: divergent questions are significantly more productive of discussion than convergent questions, and higher level questions gain much more response than lower level questions. The comparison of consistent versus inconsistent questions is in the predicted direction; but the results must be accepted with caution because of the marginal level of statistical reliability.

A further, more fine-grained analysis helps assess the generality of these three comparisons, and also sheds light on the weak results with regard to Hypothesis Ic. This analysis involves a cross-classification such that each paired comparison is examined *within* each of the subcategories of the other variables. Thus, for example, where Hypothesis Ia predicts that divergent questions will get more mileage

TABLE 2

MEAN NSS SCORES FOR THREE SETS OF VARIABLES: CONVERGENT VS. DIVERGENT,
HIGH VS. LOW LEVEL, AND CONSISTENT VS. INCONSISTENT

Variable	Mean	S.D.	N
1. Divergent	4.13	3.23	68
2. Convergent	1.67	1.16	60
Difference: 1-2	+2.46**		
3. High Level	3.90	3.95	70
4. Low Level	1.57	1.32	53
Difference: 3-4	+2.33**		
5. Consistent	3.16	3.63	94
6. Inconsistent	2.24	1.77	34
Difference: 5-6	+0.92*		

NOTE: The N for categories 3 and 4 combined is smaller than the other groupings because five questions had so little content that they were unclassifiable as to level. They are discussed later as "unfocused" questions.

* $P < .10$ (t-test, one-tailed)

** $P < .001$ (t-test, one tailed)

than convergent questions, we ask: is the direction of difference the same within the multiple question subgroup as it is within the single question subgroup? And is it the same within the high level subgroup as it is within the low level subgroup? In other words, do the results run consistently through the data, or are there reversals?

The general answer is that the patterns are quite consistent throughout the data. There are 12 direction-of-difference comparisons to be made using the cross-classification system just described, and if we do this separately for Samples 1 and 2, there are a total of 24. Of these, 22 are in the directions predicted by Hypotheses Ia-c. While the individual cells are not statistically significant due to the small N's involved, this shows that all three of the variables do have consistent effects. There are two reversals, however, such that in each case a direction in Sample 1 is reversed by that in Sample 2. Both of these are in the convergent subcategory, and the result is to weaken the differences involved in testing Hypotheses IIb and IIc. Among the divergent questions, however, both hypotheses are well supported, as Table 3 shows.

While the reasons for this variation in pattern are not wholly clear, it is possible to speculate that the convergent versus divergent variable sets a *context* for the operation of the other qualities. Perhaps asking a high level question in a "one answer" frame inhibits the richness of response that its intellectual complexity implies, and thus overshadows the differential impact of high versus low level

TABLE 3

COMPARISON OF MEAN NSS SCORES FOR TWO PAIRS OF VARIABLES (HIGH VS. LOW INTELLECTUAL LEVEL, AND CONSISTENT VS. INCONSISTENT) WITHIN THE DIVERGENT SUBCATEGORY

Variable	Mean	S.D.	N
1. High Level	4.68	4.38	50
2. Low Level	1.92	1.73	13
Difference: 1-2	+2.95**		
3. Consistent	4.70	4.38	47
4. Inconsistent	2.74	1.35	19
Difference 3-4	+1.96*		

* $P < .05$ (t-test, one-tailed)

** $P < .025$ (t-test, one-tailed)

structuring. And it may be that the confusion engendered by an inconsistent question is kept in check by the specificity provided by a convergent focus. Conversely, the openness of the convergent questions would allow more free play for the other factors to exert influence on discussion response.

Another type of analysis helps clarify the picture with respect to Hypothesis IIc. Although the consistent questions are presented as one group in Tables 2 and 3, this group is actually composed of two subtypes: single questions, consisting of only one sentence; and multiple questions, which contain two or more repetitions or paraphrases of the same question. Thus there are two variables, not one, involved: the consistent versus inconsistent distinction referred to in Hypothesis IIc, and the distinction between single-sentence and multiple-sentence questions. We have, as a result, three subtypes of questions; there are, of course, no single-inconsistent questions since there must be more than one question-sentence for inconsistency to arise.

By comparing the multiple-consistent and multiple-inconsistent questions, we can control for the single versus multiple variable, and thus provide a clearer test of Hypothesis IIc. In contrast to the data in Table 2, the difference here is quite sharp. The mean for multiple-consistent questions is 3.69 (S.D. = 2.67, $N = 35$), which is significantly higher ($P = .01$ by t-test) than the mean of 2.24 presented in Table 2 for (multiple) inconsistent questions. The mean NSS score for single-consistent questions is between the other two groups at 2.89 (S.D. = 2.52, $N = 56$). It is higher than that of the multiple-inconsistent group, and the statistical significance level is about .10. Thus, by breaking the consistent group into its two com-

ponents, we produce one clearly reliable difference and one which is at the same marginal level as was the consistent group as a whole. This occurs because of the markedly lower variances of the separate subgroups.

The two consistent subgroups are also different from each other, again at a marginal level of statistical significance ($P < .10$). This leaves us with something of a puzzle: why should single questions be lower in mileage, even marginally, than multiple consistent questions? If clarity and directness are positive qualities, as Hypothesis IIc asserts, then we might expect an advantage in favor of single questions. Since the consistency variable is controlled in the single-consistent versus multiple-consistent comparison, it seems evident that the multiplicity or redundancy factor must be understood in some other way. I will return to this point after discussing the data related to Hypothesis IIIc, which show a very similar patterning of results.

Finally, Hypothesis IID states that questions with built-in structure will generate more discussion than unfocused questions. In testing this hypothesis, a consideration arises that is related to the cross-classification discussed previously. The question is: what is the appropriate group with which to compare unfocused questions? They tend to be "anything goes" questions with very little specific content; thus they are plainly divergent (even to the point of absurdity!), but the lack of content makes it difficult to designate an intellectual level at all. It therefore seems reasonable to compare them with all other divergent questions at both high and low thinking levels. A further point is that all the unfocused questions happen to be consistent in content; hence, the test of Hypothesis IID is made within the consistent subgroup only.

As Table 4 shows, the comparison with high level, structured, divergent questions is in the predicted direction: the mean NSS

TABLE 4
MEAN NSS SCORES FOR THREE CATEGORIES OF DIVERGENT, CONSISTENT QUESTIONS:
STRUCTURED HIGH LEVEL, STRUCTURED LOW LEVEL, AND UNFOCUSED

Variable	Mean	S.D.	N
1. Structured High Level	4.97	4.76	37
2. Structured Low Level	2.17	1.82	12
3. Unfocused	2.60	1.74	5
Difference: 1-3	+2.37**		
Difference: 2-3	-0.43*		

* N.S. (t-test, one-tailed)

** $P < .15$ (t-test, one-tailed)

score for structured, high level questions is 4.97 versus 2.60 for unfocused questions. This is marginally significant statistically, and the finding is made more questionable by the low N for unfocused questions. Also, there is no significant difference when unfocused questions are compared with structured, low level divergent questions. Thus, on balance, Hypothesis II_d receives qualified support. The strongest statement that can be made is that as a strategy for eliciting open-ended, analytical discussions, the unfocused question is noticeably less effective than the structured divergent question. Of course, a substantially larger pool of unfocused questions would be necessary in order to draw truly solid conclusions on this point.

Hypothesis III concerns the relationships between question complexity and wait-time.⁷ Hypothesis III_a predicts that students will use more wait-time in responding to higher-level as opposed to lower-level questions; however, the data do not support this assumption (see Table 5). High level questions had a slightly longer mean

TABLE 5
MEAN WAIT-TIME SCORES FOR THREE SETS OF VARIABLES: HIGH VS. LOW LEVEL, SINGLE VS. MULTIPLE CONSISTENT VS. MULTIPLE INCONSISTENT, AND STRUCTURED VS. UNFOCUSED

Variable	Mean	S.D.	N
1. High Level	1.68	2.16	60
2. Low Level	1.58	1.41	35
Difference: 1-2	+0.10*		
3. Single-Consistent	1.60	1.89	41
4. Multiple-Consistent	0.82	0.89	33
5. All Consistent (3 & 4)	1.25	1.58	74
6. Multiple-Inconsistent	2.50	2.59	21
Difference: 3-4	+0.78***		
Difference: 3-6	-0.90**		
Difference: 4-6	-1.68****		
Difference: 5-6	-1.25***		
7. Consistent, Divergent, Structured, High Level	1.44	1.80	30
8. Consistent, Divergent, Structured, Low Level	1.51	1.48	8
9. Consistent, Divergent, Unfocused	3.23	0.96	4
Difference: 7-9	-1.79***		
Difference: 8-9	-1.72***		

* N.S. (t-test, one-tailed)

** P < .10 (t-test, one-tailed)

*** P < .025 (t-test, one-tailed)

**** P < .01 (t-test, one-tailed)

⁷ N's in the section dealing with Hypothesis III are somewhat smaller than elsewhere because one of the 7 videotapes was inadvertently erased prior to this aspect of data analysis.

wait-time than low level questions, but the difference does not even approach statistical significance. Moreover, this finding is mirrored when the data are broken down into subcategories defined by the other variables (convergent versus divergent and consistent versus inconsistent). All differences are small and inconsistent in direction.

Hypothesis IIIb compares structured and unfocused questions, predicting longer wait-times for the latter. The approach taken in testing this hypothesis follows the same line of reasoning described in connection with the test of Hypothesis IIb: that since unfocused questions are divergent and ambiguous with respect to intellectual level, they should be compared with structured divergent questions at both levels of complexity; and that since all the unfocused questions in the sample were consistent in content, the comparison group should be made up of consistent questions only. The results of both these comparisons support Hypothesis IIIb. As Table 5 shows, the shortest mean wait-time occurs with the structured, high-level questions, and the structured low-level questions are not too dissimilar. Furthermore, both comparisons are statistically significant, although the small number of unfocused questions must make this conclusion a cautious one.

Finally, the data in Table 5 test Hypothesis IIIc, which states that inconsistent questions will be associated with longer wait-times. Like the NSS scores, the wait-time data show marked differences among single, multiple-consistent, and multiple-inconsistent questions, and so these three groupings are analyzed separately. From this it is clear that the inconsistent group has the longest mean wait-time (significantly different from the other two), which supports Hypothesis IIIc.⁸ Moreover, the ordering of groups is constant across Samples 1 and 2, and when examined separately within the convergent, divergent and high level subcategories. There is a reversal of order between the single and inconsistent groups among low level questions, but by and large it can be said that the pattern is a general one.

There is still, however, the puzzling finding that single-consistent questions have noticeably longer wait-times than multiple-consistent questions, to a statistically significant degree. This mirrors the pat-

⁸ This analysis includes only question-types which are represented in all of the three categories of Table 5. Unfocused questions ($N = 4$) and lower-level divergent questions ($N = 6$) are not included.

tern found with NSS scores (Hypothesis IIc), and such a similarity across two quite different indices of student response is worth some thought. It has led me to an ambiguity in the concept of wait-time, one which leads in turn to a broader and more productive view of question response.

If we treat "wait-time" as literally that, an empty period in which the teacher awaits a reply, then it is reasonable to simply count the number of seconds between the end of the question and the beginning of the first response. But, as already mentioned, this empty space must be filled with considerable information processing before a reply can be made; "wait-time" is really "think-time" from the student perspective. This point bears on the single versus multiple distinction in an important way: for while wait-time begins when the teacher stops talking, *think-time begins after the first complete statement of the question*. During the repetitions or paraphrases used in multiple-consistent questions, the student can be checking his or her understanding of the question, enriching his or her appreciation of its meaning, and formulating a response. In other words, the answer is being prepared while the instructor is still speaking, and it is really not surprising that students leap in to talk so quickly when the opportunity arises. By contrast, of course, single questions do not embody any think-time (other than wait-time); and multiple-inconsistent questions complicate its function by adding new questions which the student must digest and try to reconcile with the initial question.

Looked at in this way, the wait-time ranking of the three groups makes a good deal of sense. This further reinforces the assumption underlying Hypothesis III: that building in appropriate time for information processing is an important ingredient in creating good discussions. And it seems likely that the same or related factors could influence the extent of response as well—not as markedly as with wait-time, but enough to account for the marginal NSS score superiority of multiple-consistent as opposed to single questions. For example, having two or three alternate versions of the question to respond to may draw in more students; the opportunity for unpressured think-time during the question may stimulate greater response; and the positive climate created by the lack of a post-question pause may encourage hesitant individuals to speak up.

The data pertinent to Hypothesis III make the assumption of a

relationship between question complexity and wait-time seem a useful one. However, there is no evidence that high level thinking *per se* requires more information processing time. Well-formed questions at any level of thinking seem to be answered quickly; it is the poorly formed questions (inconsistent and unfocused) that require the most struggle from students.

A final point is of interest in connection with wait-time. Since this, like the NSS score, is a measure of student response, what is the relationship between the two? Unfortunately, the picture is far from clear. Taking the sample as a whole, the correlation is zero-order ($r = .11$), but the subcategory patterns vary widely. Among high-level questions (both convergent and divergent), the correlations are zero-order or negative; that is, the shorter the wait-time the greater the number of student responses. Among low-level questions, however, the correlations tend to be moderately positive, which means that the long wait-time questions get the most response. And because the correlations are generally modest (median $r = \pm .26$) and the subcategory sample sizes small, all of the relationships fall short of statistical significance. Thus, there is no clear-cut evidence that "silence is failure." More research and analysis are needed to clarify this issue.

Exploring Hypothesis IV involved inductive, impressionistic work which would yield themes or patterns running through the data. Sample 1a was used to generate specific categories and hypotheses, which were re-checked on Sample 1b and formally cross-validated on Sample 2. Rather than further describe this "discovery" process in abstract terms, I would like to invite the reader to participate in it directly via a brief exercise (see Table 6) which we use to introduce the classification system in instructional development workshops. This exercise, while not covering all the categories which emerged, closely approximates the complete coding system.

In identifying question matches, the reader will have followed his or her own inductive process, which is probably similar in approach to mine but may have yielded somewhat different results.⁹ The formal classification procedure is partly based on previously identified variables: convergent versus divergent, higher versus lower level, multiple versus straightforward, and vague versus structured. Other distinctions are also introduced.

⁹ The exercise, is, of course, set up to highlight the research categories, and in practice, workshop participants tend to recreate them about 90% of the time.

TABLE 6
QUESTION MATCHING EXERCISE

Instructions:

The 14 questions below consist of seven pairs of similar questions in mixed order. The purpose of this exercise is to heighten awareness of the characteristics of various sorts of questions through noting similarities and differences. Please compare the questions in the two columns below, in terms of their wording and their likely effects on discussion. When you identify two that seem similar, indicate this by putting the number from the question in the left-hand column in the blank next to the right-hand matching question.

- | | |
|---|--|
| 1. So where is this wild boy better off?
In the forest where he started, or in civilization being socialized? | —A. (Instructor reads a sentence from the novel under study): Well, that's a very rich sentence . . . there's a lot there . . . OK, what's there? |
| 2. So in this story, when's the point of truth for Kurtz? | —B. Any comments on Plato? |
| 3. What about the lecture? | —C. He talks about envying one character. Who was it? |
| 4. What possibilities are there for refuge in <i>A Farewell to Arms</i> ? | —D. How do you interpret what the narrator tells you about the hero? What do you make of his return from law school? Why did he decide he didn't really expect too much? |
| 5. So, we're talking about the fact that everybody's roles are changing, how—we've mentioned religion and education, how did religion and education during this period affect these changes, or how did the changes affect the kind of religion and education people had? . . . Let's start with religion . . . have women always had a sort of divine place in religion? | —E. Is Ivan Illych a victim of society, or did he create his problems by his own choices? |
| 6. Let's see if we can make any generalizations about the play as a whole, from the nature of the opening lines. | —F. What was the most important reason for the revolution's failure? |
| 7. What was the name of that institution? | —G. What kinds of things is Hamlet questioning? Not just in his soliloquy, but broadly, throughout the whole play? |

In general, the more fruitful question types are *divergent*, *higher level*, *straightforward*, and *structured*, a finding that is consistent with the individual-variable results used to test Hypothesis II. However, the use of a "species" typology also unearthed some more subtle variations on these themes, yielding results which are particularly useful in designing teaching strategies. The types are presented below in approximate descending order of "mileage" scores. Accompanying them are the overall mileage ratings and the exercise items which correspond to the category in question. Following the "species" descriptions is a table giving cross-validation and summary statistics for the entire category system.

1. *The Playground Question* (NSS = 5.08)¹⁰

6. Let's see if we can make any generalizations about the play as a whole, from the nature of the opening lines.
- A. (Instructor reads a sentence from the novel under study): Well, that's a very rich sentence . . . there's a lot there . . . OK, what's there?

This type is so named because the instructor designates a specific intellectual sphere (the "playground") for discussion, and then gives students the widest possible latitude in approaching it. In effect, he or she is saying, "Here's a playground that should be interesting; let's all stay here so we can do something in common, but you go ahead and choose the games." This "playground" may be a poem, a character in a play, a philosophical concept, "the opening lines," a sentence, etc., and is circumscribed quite carefully in light of the instructor's teaching goals. The hallmark of the Playground Question is the invitation to explore, which is typically offered with such phrases as "How do you interpret . . . ?", "What can you draw from . . . ?", "What are the possible meanings of . . . ?", and the like. This sort of invitation leaves open the question of what concept, category, or theme the student will use to make sense of the raw material offered.

2. *The Brainstorm Question* (NSS = 4.88)

4. What possibilities are there for refuge in *A Farewell to Arms*?
- G. What kinds of things is Hamlet questioning? Not just in his soliloquy, but broadly, throughout the whole play?

This name is borrowed from the brainstorming technique, in which any and all ideas or solutions are sought in response to a specific question or problem. In contrast to the Playground Question, the primary structure is *thematic*, with the subject matter being less tightly focused. Thus, in the two examples from the exercise, the entire novel or play is within the scope of the question, but the issue to be addressed is quite specific ("refuge" in one instance, and "questioning" in the other).

¹⁰ As indicated earlier, the NSS score refers to the average number of separate student comments following questions falling into a given category.

3. *The Focal Question* (NSS = 4.29)

1. So where is this wild boy better off? In the forest where he started, or in civilization being socialized?
- E. Is Ivan Illych a victim of society, or did he create his problems by his own choices?

The Focal Question is oriented around an issue which calls for decisions. The instructor poses a limited number of alternatives (usually two or three, and rarely more than five) and asks students to take stands which they are to justify during the discussion. These stands call for higher order thinking and require students to marshal various kinds of information to support their views. There is a tone of debate or persuasion generated by Focal Questions; while the groups as a whole may or may not reach a single conclusion, the effort to exert influence toward this end affects the shape of the discussion. Thus, the question's structure comes from the alternatives posed, while openness stems from the fact that students may draw on a wide range of information to support their views. For example, in the "Wild Child" film discussion (Exercise Question 1), students could (and did) bring in everything from ethics to developmental psychology to back up their positions.

The three questions discussed so far, the Playground, Focal and Brainstorm Questions, are clearly the most productive quantitatively and share characteristics which suggest the label "Structured Divergent Questions." Each in its own way provides a clear focus which prevents confusion and wandering (a delimited chunk of the material, a thematic thread, or a set of debatable alternatives); at the same time they encourage students to express a variety of ideas and to marshal many kinds of information in order to pursue the discussion. By contrast, the question types to be described below seem to miss this synthesis, by being either too open and vague or too narrow and mechanical.

4. *The General Invitation* (NSS = 2.60)

3. What about the lecture?
- B. Any comments on Plato?

This question is a Playground Question that has lost its bound-

aries. Since the definition of the playground is the chief source of structure in such questions, it is crucial that this definition be clear and specific; it should be carved out of the material on the basis of a rationale as to what will stimulate student learning. By contrast, the “playground” in a General Invitation is broad, vague, and often defined by a handy or conventionally obvious unit of the course such as “the lecture,” “the book,” “Plato” and so on. In this context—or lack of context—the open invitation becomes simply license to wander all over the place: a starting point for the classically formless or unfocused discussion. Such questions can elicit response when students are able to provide some direction themselves—when they are well prepared, interested in the topic, and/or assertive. But students are probably as apt to freeze in confusion as they are to use the freedom offered them. Because of these problems, it is understandable that General Invitations only generate about half the mileage of Structured Divergent Questions.

5. *The Lower-Level Divergent Question* (NSS = 1.92)¹¹

These questions have the outward form of Structured Divergent Questions, but unlike them deal with the lower levels of Bloom’s taxonomy; they call for memory or comprehension answers. For example:

- a. Were farmers richer or poorer than city dwellers in this period?
- b. What are the names of some other generals during that war?

Example a has the form of a Focal Question, involving alternatives, while example b is structured like a Brainstorm Question. Although the divergent nature of these questions will sometimes produce a high NSS score, the lack of content richness limits their potential; the NSS score is almost the same as that for the convergent questions to be discussed below.

6. *The Analytic Convergent Question* (NSS = 1.95)

2. So in this story, when’s the point of truth for Kurtz?
- F. What was the most important reason for the revolution’s failure?

These questions are distinguished by the fact that a single, cor-

¹¹ This category is not included in the exercise.

rect answer is implied. Some analysis or inference is necessary, but examination of the relevant material will lead straightforwardly to the conclusion. Thus, both exercise examples emphasize, by their phrasing ("*the* point of truth," "*the most important* reason") that there is a single end-point which the instructor has in mind and which the students can be expected to identify. Such questions utilize application or analysis-level thinking within a convergent context. This involves comparisons, cause-and-effect statements, drawing of straightforward inferences from limited material, and other similar operations.

7. *The Quiz Show Question* (NSS = 1.45)

7. What was the name of that institution?
C. He talks about envying one character. Who was it?

While they may pay off handsomely on daytime television, these questions create an impoverished intellectual atmosphere. They are convergent and call for memory or comprehension level responses: facts, information, definitions, descriptions of events, and the like. The responses to them are generally brief, and they have the lowest overall NSS score in the category list.

In addition to the seven principal categories described above, one of the following four categories was applied to all questions. These codings, derivations of the multiple versus single and consistent versus inconsistent distinctions, are independent of which category of the first seven has been assigned.

8. *The Single Question* (NSS = 2.89)

These are straightforward questions which involve only one question-sentence; ambiguity is reduced to a minimum. All exercise questions except 5 and D are of this type.

9. *The Multiple Consistent Question* (NSS = 3.69)¹²

These questions are multiple, in that they contain more than one question-sentence. However, they are consistent in that the content of the question, and the level of thinking required to deal with the question, remains approximately constant throughout. Thus, in essence, only one question is being posed to students, even though they

¹² This category is not included in the exercise.

must absorb two or more versions of the question before responding. The data indicate that these questions function somewhat better than Single Questions in terms of student response. For example:

Well, does Kafka like religion? . . . Is our impression that Kafka's favorable to the development of Christianity? Are we meant to clap our hands and praise Christianity after reading this story?

10. *The Shotgun Question* (NSS = 2.50)

5. So, we're talking about the fact that everybody's roles are changing, how—we've mentioned religion and education, how did religion and education during this period affect these changes, or how did the changes affect the kind of religion and education people had? . . . Let's start with religion . . . have women always had a sort of divine place in religion?
- D. How do you interpret what the narrator tells you about the hero? What do you make of his return from Law school? Why did he decide he didn't really expect too much?

The Shotgun Question is multiple, in the sense of containing more than one question-sentence, and inconsistent. It may contain two or more separate content areas; it may embody multiple levels of thinking, as when an instructor simultaneously asks for complex analysis and factual information; or it may involve both of these inconsistencies. We call these Shotgun Questions because they often seem to be fired off in the hope that at least one fragment will hit the target. The instructor may be hoping to provide "something for everyone," be insecure about the effectiveness of any given question, or simply be thinking aloud between the beginning and the end of the question. The result is often self-defeating, because students are puzzled about what is expected; they must sort through the various questions being asked, select among them, and/or reconcile incompatibilities before responding. This is apt to heighten anxiety and inhibit participation.

11. *The Funnel Question* (NSS = 2.18)

(Note: Question #5 under "The Shotgun Question" is also a Funnel Question.)

The Funnel Question is a Shotgun Question with some particular characteristics. It is consistent as far as topic is concerned, but inconsistent as to level of thinking; and the level of thinking moves

down the hierarchy of complexity. Thus it "funnels" students from a general playground into the narrow chute of a convergent question. Sometimes these questions tumble out all at once, and at other times there is a notable pause between each subquestion. In either case, it is as though the instructor began with a broad question—a General Invitation or a Structured Divergent Question—and, fearing that it would fall flat, rushed in with progressively more and more structure in hopes of pulling a response from students. At times one has the impression that the opening, general phase of a Funnel Question is really "fishing": the instructor has a quite specific answer in mind but hopes that students will find it on their own. Then, when this fails to occur, he or she keeps giving hints and eventually arrives at a convergent reformulation which conveys quite clearly the sort of response originally desired. Funnel Questions are about as effective as Shotgun Questions in eliciting student response: about two-thirds as efficient as Single or Multiple Consistent Questions.

Table 7 displays the question typology data in a form amenable to cross-validation. The pattern of mileage scores is quite similar when Samples 1 and 2 are compared, both in terms of absolute means and in terms of the ordering of question types. The latter point can be stated as a hypothesis: if there are consistent differences among question types in the degree of student response, the rank-

TABLE 7
MEAN NSS SCORES FOR ALL QUESTION TYPES

Question Type	Samples 1 & 2			Sample 2			Combined Sample		
	Mean	S.D.	N	Mean	S.D.	N	Mean	S.D.	N
<i>Major Question Categories</i>									
Playground Question	5.00	2.73	7	5.17	4.37	6	5.08	3.59	13
Brainstorm Question	5.57	7.41	7	4.33	3.13	9	4.88	5.47	16
Focal Question	4.43	3.94	14	4.00	3.59	7	4.29	3.83	21
General Invitation Question	2.75	3.83	4	2.00	0.00	1	2.60	3.90	5
Analytic Convergent Question	1.38	0.70	8	2.33	1.31	12	1.95	1.20	20
Low Level Divergent Question	1.67	1.08	6	2.14	1.61	7	1.92	1.41	13
Quiz Show Question	1.50	1.51	18	1.41	0.78	22	1.45	1.17	40
<i>Supplementary Question Categories</i>									
Multiple Consistent Question	3.30	3.60	23	4.42	3.58	12	3.69	2.67	35
Single Question	3.30	4.54	26	2.53	2.59	30	2.89	2.52	56
Shotgun Question	2.57	1.55	7	2.43	1.68	7	2.50	1.14	14
Funnel Question	2.22	1.83	9	2.13	2.03	8	2.18	1.36	17

ordering of the types should be the same or similar across samples. This is tested by computing a rank-order correlation (ρ) between the rankings generated from the two sample groups. For the seven primary question types, $\rho = +.82$, $P < .025$. Among the four secondary categories, $\rho = +.95$, although the N of 4 is too small to satisfy the assumptions of the significance test. And for the total of 11 categories, $\rho = +.78$, $P < .01$. Thus, the result is clearly positive.

A second hypothesis generated by the inductive process of identifying question types is that Structured Divergent Questions will lead to more productive student discussion than other types. This conclusion is overwhelmingly clear from the data in Table 7, which show that the former have a mean mileage rating nearly three times as great as the latter. This conclusion, generated from Sample 1, is supported in Sample 2 ($P < .01$ for both). And the t -test for the combined samples is highly significant, with $P < .00001$.

The data presented thus far provide an affirmative answer to our original, general question: does the question's verbal structure *per se* affect the nature of student response? Beyond this, however, it is useful to know whether the relationships between question type and mileage score hold for all instructors or only certain ones. A finer-grained analysis shows that, within the limits of our seven-instructor sample, the findings are quite uniform. The data supporting this conclusion are obtained by looking at the direction of difference between the pairs of variables used to test Hypotheses Ia, Ib, and Ic, separately for each instructor. This yields 7×3 or 21 paired comparisons,¹³ all of which should be in the predicted direction if the data are internally consistent. In fact, 20 of the 21 differences are as predicted, and in the one instance of reversal the two means are not particularly far apart. While, of course, most of these differences are not statistically significant due to the small N 's involved, the 20:1 ratio is in itself significant by sign test ($P < .001$). Thus, we can assert with some confidence that *any* instructor (within the population represented by our sample group) will get good results from using the high-mileage questions.

But how varied are these instructors? Is it possible that teachers who are closer to the extremes of style or competence might show

¹³ The structured versus unfocused distinction must be left out of consideration here because of the rarity of the latter question type.

different results? We have no objective index of overall competence, but my impression is that the range is fairly narrow: all our instructors are reasonably competent but not outstandingly skillful. And it may be, for example, that a very inept instructor would get little discussion from even our best questions because of an inhibiting general climate or poor followup.

With respect to teaching style, however, there is some useful evidence. This is generated by measuring instructors' predilections for using certain of our question types rather than others. Specifically, what percentage of the time does a given teacher use divergent as opposed to convergent questions? High or low level? Consistent or inconsistent? These three ratios denote style variables which can be used to assess the consistency of instructor behavior and to provide a context for other discussion features.

To establish whether these are truly "styles" rather than random clusterings, it is first necessary to assess the degree of continuity from Sample 1 to Sample 2. For % divergent (%D) and % high level (%H), the rankings of the seven instructors were very similar in the two samples. The Spearman rank-order correlation for %D is $+ .79$, $P < .03$, and for %H it is $+ .96$, $P < .02$. However, % convergent (%C) was much less uniform, the correlation being only $+ .31$, which is not statistically significant. This variable does not appear to qualify as a consistent style factor.

A further point of note is that the %D and %H scores are highly similar, due primarily to the fact that all the Structured Divergent Questions are counted in both categories. The rank-order correlation between the two variables (for both samples combined) is $+ .96$, $P < .02$. Thus, in effect, we have identified a single style factor with several aspects. For this reason, the analysis below is pursued using %D only; quite similar results would be obtained by using %H. %D is the more useful of the two because the range of scores is quite wide, indicating that our sample represents nearly the full range of style possibilities: from 100%D at one extreme to 28% D at the other.

With the foregoing context established, we can now ask: does instructor style affect question mileage? Specifically, does an instructor whose style is predominantly divergent get more mileage overall as compared to one who usually asks convergent questions? The answer is "yes," but this is a contaminated finding; since diver-

gent questions are more effective, it follows mathematically that one who uses a greater proportion of them will have a higher overall NSS score.

This distortion can be controlled by examining the relationship between %D and NSS *within* the divergent and convergent subcategories taken separately. Here, the results are quite clear. Among our seven instructors, the higher the %D score, the greater the question mileage (mean NSS score). This is most striking when we consider only divergent questions: the rank-order correlation between %D and NSS is $+.82$, $P < .03$. Among the convergent questions, the relationship is still present, but weaker: $\rho = +.60$, $P < .15$. Moreover, these mileage differences are far from trivial, especially among the divergent questions; the highest scoring instructor had a mean NSS score of 5.71, while the score of the lowest was 1.43. The range for convergent questions was 3.33 to 0.94.¹⁴

The most plausible interpretation of this finding is that when the overall style is divergent, students develop expectations which make them more ready to respond; probably they gain practice in dealing with such questions and feel more expansive and free from judgments. It is interesting that this climate factor apparently carries over even when convergent questions are asked; perhaps in divergent-oriented classes students give ambiguous questions the benefit of the doubt, interpret even a convergent structure in a divergent way, or feel freer to raise divergent issues around overtly convergent questions.

To summarize the findings: the verbal form of questions does influence the extent of discussion response, and it is possible to identify several types of questions which have the most productive results. These, collectively called Structured Divergent Questions, have the common characteristic of providing students considerable freedom of expression within the frame of a definite focus which holds the class together and provides a sense of continuity and direction. This impact is predictable and consistent cross instructors. However, it occurs within a larger context of instructor style: although all seven instructors got better mileage when using such questions, the most fruitful results seem to come from building a

¹⁴ It should be borne in mind that these instructor differences do not contradict the finding that for *all* instructors divergent questions were superior to convergent questions in NSS score; the evidence presented here concerns the *magnitude* of that superiority.

consistent style around the structured divergent approach. Finally, the wait-time findings show the importance of allowing students adequate time for information-processing. And they provide a further reason for avoiding certain types of questions: those which generate too much ambiguity and confusion, and thereby produce long, discussion-chilling, post-question silences.

IMPLICATIONS FOR BETTER TEACHING

These data begin to establish a firmer footing for teaching improvement recommendations. Some key points are:

1) It is probably better to avoid ambiguity about convergence or divergence. Ask questions in such a way that you communicate *clearly* that multiple answers are acceptable or discussible, and follow this up with prompting and encouragement which invites second and third responses to the same question.

2) Familiarize yourself with Bloom's (1956) taxonomy and sprinkle your discussions liberally with questions requiring higher levels of thought: analysis, synthesis and evaluation.

3) Don't let complexity run rampant. The data show us two forms of cognitive overload, the Shotgun and Funnel Questions, which seem to produce confusion and withdrawal. In addition to the "brain cell torque" (to borrow a student phrase) needed to grasp the question, search one's storage-retrieval system and verbalize an answer, students must sort through mixed messages and overcome the fear of seeming foolish by answering the wrong question. This leads to poorer mileage and longer wait-times; so state your question crisply and take your chances with it.

4) Bear in mind that students may read your question in terms of the kind of thinking expected, and that it is confusing and inhibiting to switch levels drastically. One reason for this may be motivational as well as cognitive: people who are at one moment confronted with a challenging, adult issue, and at the next with a Micky Mouse, "When was the war of 1812?", type of question may oscillate psychologically, advancing a few years at one moment and regressing to high school the next. For students at a late-adolescent transition point, whose adult identities are still very fluid (Erikson, 1960), this process can be unsettling.

5) Don't be seduced by "pure" freedom. Many instructors, when they first decide to break away from the rigidity of low-level convergent questions, swing to the other extreme and give students a gigantic, formless playground in which to get lost. General Invitations do produce a bit more interaction than Quiz Show Questions, but they are not consistently effective. Probably they are most likely to succeed

when students are (a) very interested in the material, (b) well prepared, and/or (c) mature and assertive. But they rarely offer anything that Structured Divergent Questions don't, so that the latter are almost always to be preferred.

6) Decide what style you want to use and be consistent with it, at least within a given class meeting. Just as mixed-level questions are less successful, so is a mixed style; high-level, divergent questions get more mileage from instructors who use them frequently and consistently. There are probably several ingredients in this. Students will tune in to the instructor's general expectations and thus be more ready to respond; they will learn to recognize complex questions and become more skilled in handling them; and they will assimilate into their own attitudes the freedom implied by this approach, and thus be more verbal in general.

7) Use only the amount of divergence you are truly willing to live with. There is nothing more inhibiting than a wide-open question which is followed by "well, that's not exactly what I had in mind" or a veiled version of the same message (communicated perhaps via a Funnel Question). Divergent questions which are in reality attempts to fish for preconceived answers will sooner or later become self-defeating and sterile.

8) The wait-time data lead to some modifications of suggestions I have made to instructors in the past. Long wait-times do *not* seem to be natural accompaniments of complex, challenging questions, but rather of ineffective questions (General Invitations, Shotgun Questions and Funnel Questions). Thus, if you are uncomfortable with the length of wait-time, this provides an additional reason to aim for Structured Divergent Questions.

9) The Multiple Consistent Question is a useful tool, in that it combines conceptual clarity with the opportunity to chew over the question and see different sides of it while the instructor is still talking. Thus, you may use redundancy as a deliberate means of shortening wait-time by building in earlier think-time. The evidence suggests that you will also gain somewhat in mileage by phrasing your questions in this way.

10) It is important also to avoid turning good questions into Shotgun or Funnel Questions by too hastily reacting to silence with a new or altered question. A much better tactic is to inquire whether the silence reflects a need for clarification or restatement; this, in effect, turns a Single Question into a Multiple Consistent Question through the use of student feedback.

11) It is important to keep in mind that some silence is a natural part of the discussion, and to use it rather than fearing it. And there is no clear evidence that "silence is failure" where well-formed questions are concerned. While most of our questions (88%) were answered within three seconds, we have examples of longer wait-times

which were followed by very productive discussions. The record is currently held by a T.A. who waited 29 seconds following a Playground Question; this turned into a discussion with an NSS score of 17, and thus was well worth the wait. Allowing for adequate think-time may also help students learn not to back away from ambiguity and intellectual frustration, but rather to stick with an initially intimidating issue until some clarity is reached.

12) Assimilate the formal properties of our three high-mileage questions into your discussion style, and use them frequently. With practice, your ability to instinctively cast subject matter into these forms should improve.

Perhaps the most important quality to grasp is a subtle blend of structure and freedom which gives a discussion momentum and yet does not let it wander indiscriminately. By taking some thought, instructors can write much better questions than those which turned up in our transcripts. The teaching process—not to mention future research—would flow more smoothly through eliminating the abortions and odd hybrids we dredged up on occasion.¹⁵ To this end, I present below some guidelines for creating Structured Divergent Questions.

The best starting point for generating Structured Divergent Questions is the focus of structure for each question type. One should pick a focus which is (1) crucial or pivotal for the material under study, and (2) rich in implications and ramifications. Specifically:

a) For Playground Questions, choose a promising sub-aspect of the material. Examples are a key scene in a novel, certain highly condensed and symbolic lines in a poem, a turning-point battle in history, a passage delineating a philosophical concept that is central to the philosopher's position, and so on. This playground should be introduced to students with some indication of its importance, richness, etc. You can then invite them to explore its internal structure and wider implications. This invitation should be issued in the open form already referred to: "What can you draw from . . . ?", "What do you make of . . . ?", "What are the implications of . . . ?", and so on.

b) For Brainstorm Questions, selecting the thematic focus is the key. This is not unlike what a literary critic, for example, might choose as a unifying theme which is refracted in various ways throughout a work. Again, breadth of scope and richness of association are important. At the same time, you must be willing to entertain surprises: new, even

¹⁵ The irony of a researcher who complains of the messiness of his subject matter turns on the fact of human self-reflexiveness. Only we humans are capable of comprehending research on ourselves and guiding our behavior accordingly.

improbable-looking connections between the theme and other aspects of the material. One of the virtues of this thematic approach is that it helps students unify a topic, book, or other unit of material, as the theme suggests interrelationships between apparently disparate subparts. Once having selected the theme, introduce it to students with some variant of the following: "This theme (issue) arises in lots of ways when we examine _____. What are some ways you can think of?"

c) Focal Questions involve either-or decisions or controversies, and so the first choice of structure is to set these terms. The alternatives should all be tenable or defensible. They should be pivotal in that one's choice among them has significant ramifications for one's view of a larger topic or sphere of discussion, and they should enable the student to draw on a sizable portion of the material in this topic area in support of one or another viewpoint. The question may involve a complex problem with several possible solutions or interpretations; a controversy (concerning values or analytic conclusions) with several possible stands to be taken; a phenomenon which can be interpreted according to two or more theories; or a complex, provocative statement which can be agreed or disagreed with.

A good process for generating all three types of questions is brainstorming, either for yourself or with the class. When the latter is done, students become acquainted with the question form, which helps them participate effectively. They can use their own interests as criteria of selection, and everyone will feel more investment and responsibility for the ensuing discussion. For example, one class concerned with ecology and environment chose as a Focal Question to discuss whether Eskimos should be restricted in their killing of seals. They recognized this as internally dynamic because it pitted against each other two values which this group of students supported strongly: the preservation of the Eskimos' independence and traditional way of life on the one hand, and the protection of an endangered animal species on the other. And they could draw widely in search of support for one argument or the other: on ethics, political decision-making and the role of government, ecological biology, anthropology, and so on.

Structured Divergent Questions can be used in many formats. They can be introduced *ad hoc* in discussion; assigned as study questions or as pre-planned discussion questions; and used as essay exam questions. Learning is strengthened when the type of thought encouraged in discussions is mirrored in the evaluation procedure as

well. Such questions can also be a powerful stimulus to learning factual information, because the latter is seen as useful for dealing with meaty problems and issues. One approach is to pose a question near the end of one class session, elicit enough discussion to catch students' interest, and then ask them to think about it during their studying in preparation for the next meeting.

The three types of Structured Divergent Questions can be expanded into classroom formats useful for structuring one or more whole sessions.¹⁶ Projecting them on such a giant screen also helps clarify their characteristics and the differences among them.

a) The format corresponding to Brainstorm Questions is, not surprisingly, the brainstorming session. This process was developed some time ago (Osborne, 1963) in order to stimulate creativity by reducing the inhibiting effects of criticism and argumentation. The basic ground rule divides the session into two parts: the first is devoted to idea production; no holds are barred, and apparently zany ideas are actively encouraged while evaluative commentary is out of bounds. Generally, the leader (teacher) records all contributions visibly, on the blackboard for example. The underlying attitude is, "Why come up with two ideas and then spend the next 30 minutes arguing over them, when in the same time period we can come up with 20 ideas and then select the best among them?" The second phase of the brainstorming session is concerned with combining ideas, applying criteria, and deciding which are truly valid and useful.

b) The Focal Question can be used in any format which enables students to declare themselves on an issue and argue back and forth. One approach is to ask for a hand vote periodically. Focal Questions also lend themselves well to a debate format, and I have found a debate variant which is extremely effective in college classes. This is called by its inventor¹⁷ the "change-your-mind-debate" because, unlike traditional debates, it encourages the participants to let the discussion influence them toward new views. The room is set up with three banks of chairs (when there are two alternatives posed); one bank is for those who espouse each of the two main positions, and the third is for those who have mixed reactions or are undecided. Students are instructed that whenever they change their minds during the debate, they are to get up and move to the area which reflects their current views. Thus,

¹⁶ While we do not have enough examples of these formats to conduct a systematic study, the three which are available (one of each) are extremely effective by the discussion mileage criteria used in this research. They all have NSS scores well beyond the range of the questions analyzed in Table 7.

¹⁷ My thanks to Zachary Seech, a graduate student in the Philosophy department, University of California, San Diego, for this idea.

at all times the pattern of the room reflects the progress of the discussion, and all present are declaring themselves even if they do not verbally participate. Often the "undecided" students become informal moderators.

c) The case-study is a Playground Question writ large, because it delineates a set of unstructured raw material which students are to interpret in their own fashion. Original source materials are used in some history courses in this way. Students draw on their general learning from the course in order to make interpretations, answer questions they have posed, or draw a portrait of the events in question. Thus, as with the Playground Question, focus is given by the concrete materials; the lens of analysis is left to the individual's choice.

The three types of Structured Divergent Questions can have quite different effects on class interaction and the learning process, especially when used in the whole-class formats described above. One may want to select among them in order to attain various teaching objectives. The following section describes how each type can be expected to function.

a) *Interaction Patterns*: Focal Questions are most apt to produce student-to-student interaction (especially when used in the change-your-mind-debate), since the main line of talk is between supporters of opposing positions. By contrast, Brainstorm Questions and brainstorming sessions minimize student cross-talk while ideas are being collected; the teacher plays quite a central role by helping students formulate ideas, recording contributions, and so on. Playground Questions are intermediate between the other two types.

b) *Collaboration and Competition*: Here the contrast is between Focal Questions, which are competitively oriented, and Playground Questions, which require students to collaborate and build on each other's analyses of the material. The Brainstorm Question tends to promote an individualistic orientation but leans more toward collaboration.

c) *Cognitive Level*: In terms of Bloom's typology, all three questions are at the higher levels of thought. Within this frame, Playground Questions stimulate analytic thought, since they involve seeing relationships and interconnections within a complex body of material. Brainstorm Questions, while they may be analytic, lend themselves particularly well to synthetic, creative types of thinking, since they suspend evaluation and invite novelty. And Focal Questions, although also usable for analysis, are most suited to evaluative issues because they ask students to make choices based on articulated criteria.

d) *Mode of Incorporating Data*: The three questions differ somewhat in how they ask students to use information during the discus-

sion. Playground Questions ask students to interpret raw data, and are thus primarily inductive in emphasis. Focal Questions are more deductive, in that they deal with the implications of holding certain convictions, the testability of various conceptualizations, and so on. And finally, Brainstorm Questions do not fall into a clear-cut category. They are inductive in directing students to seek patterns in the material, but deductive in the sense that the pattern is already provided rather than discovered.

APPLICATIONS TO FACULTY DEVELOPMENT

We use these concepts at several points in teaching development programs. First, we introduce the main questioning concepts and provide opportunities to practice identifying and distinguishing the categories. This can be done through a formal presentation followed by examples, or via the inductive, "discovery" exercise presented previously.¹⁸ A second application activity involves using the complete coding scheme to analyze sample materials or the instructor's own classroom questions. Somewhat the same effect is achieved by discussing sample videotapes more informally in the light of the main concepts.

A further step is added by asking instructors to write their own Focal, Playground or Brainstorm Questions using subject matter currently relevant to them. In many cases, these practice questions find their way into study question lists, essay exam questions, and the like. Practice in using these as discussion questions is provided through workshop role-playing, in which participants serve as simulated students. And finally, instructors who want to try the questions in class may have their sessions videotaped in order to compare intentions with results.

SOME CONCLUDING THOUGHTS

The patterns uncovered in this research should make the ingredients of good teaching somewhat less mysterious. I often hear instructors talk about unexpectedly "good" or "bad" classes; when

¹⁸ A second form of the inductive exercise is called "matching." Each workshop participant is given a single question, and all are instructed to mingle by engaging in a series of one-to-one conversations in which the two individuals compare their questions and decide whether or not they are similar. In doing this, they articulate to each other various constructs for defining similarity. When the group reconvenes, these impressions are shared and a category system is constructed which usually approximates the formal one presented here. Because of the informal interaction it generates, the matching version of the exercise is a useful introductory "ice-breaker" in workshops.

the result is positive, people accept their good fortune happily enough but with little sense that they know how to make it recur. The thoughtful use of our more powerful question types should give teachers another tool for constructing successful discussions.

I find little support in our findings for the fear that rigor will be sacrificed in producing open, involved discussions (or vice versa). I suspect that this fear may stem from some half-truths about both qualities. Specifically, if one associates rigor with gaining a lot of knowledge, and thus with Quiz Show Questions or Analytic Convergent Questions, the best discussion *will* be lost. And conversely, if one associates free discussion with the shelving of standards or the substitution of opinion for the thoughtful use of information, rigor *will* diminish. The latter is most apt to happen through overuse of General Invitations with the vagueness and lack of direction which results.

The Structured Divergent Questions help us to have our cake and eat it too. They provide space for students to bring in new ideas, yet each type has its own built-in directionality. While they encourage students to think for themselves, there is also encouragement to do that thinking *about* something: to incorporate the facts of the discipline into the discussion.

Use of these questioning strategies should help students develop intellectually and become more nearly full partners in the learning process. Because they can be explicitly defined and explained to students, the Structured Divergent Questions are a good basis for giving roles in the class. They can organize their own thinking through constructing questions, leading discussions, and so on. The overall result should be a more sophisticated approach to learning in general. In this connection, it is interesting in that the three Structured Divergent Questions call for thinking characteristic of the higher intellectual and ethical stages delineated by Perry (1970) in his developmental study of college students. It may turn out that these teaching tools can help us foster such maturational progress as well as helping students acquire cognitive skills *per se*.

Finally, the conclusions reached here suggest many issues for further exploration. Are these question categories exhaustive, or will other "species" turn up? What differences would we see in studying other types of classes, for example in the natural sciences? Is the questioning pattern of the instructor related to the way stu-

dents rate him or her on evaluation questionnaires? Is questioning pattern related to instructor personality? Is it related to outcome: that is, to what students learn as measured by various testing criteria? Do students with different personality patterns, learning styles, or levels of development respond differently to our various questioning patterns? Answering these and other questions should help us construct a more complete anatomy of the effective discussion.

BIBLIOGRAPHY

- Bloom, B. S., *Taxonomy of Educational Objectives*. New York: Longmans Green, 1956.
- Erikson, E. H., "The Problem of Ego Identity." In Stein, M. R., Vidich, A. J., & White, D. M., *Identity and Anxiety*. Glencoe, Ill.: The Free Press, 1969. Pp. 37-87.
- Guilford, J. P., *Intelligence, Creativity, and Their Educational Implications*. San Diego: R. R. Knapp, 1968.
- Napell, S., "Six Common Non-Facilitating Teaching Behaviors." *Contemporary Education*, Vol. XLVII, No. 2, Winter, 1976. School of Education, Indiana State University.
- Osborne, A. F., *Applied Imagination*. New York: Scribner, 1963.
- Perry, W. G., Jr., *Forms of Intellectual and Ethical Development in the College Years: A Scheme*. New York: Holt, Rinehart & Winston, 1970.